We would like to congratulate the authors on their impressive contribution, which has already stimulated further research in the area and will certainly inspire many other developments. The idea of running parallel computations for Markov chain Monte Carlo, which is sequential by nature, has recently been attracting a lot of attention in the Bayesian community, e.g. [Scott et al., 2016], [Dai et al., 2019], [Rendell et al., 2018]. We believe the authors achieved a major breakthrough in this field by proposing a general framework for parallelising computations, applicable to many MCMC algorithms [Heng and Jacob, 2019, Middleton et al., 2018].

Moreover, an important advantage of this method is applicability to models composed of modules [Liu et al., 2009, Plummer, 2015, Zigler, 2016, Jacob et al., 2017], where due to misspecification of the full model it may be beneficial to estimate parameters of these modules sequentially, as discussed in Section 5.5. The law of total expectation ensures that the estimators of integrals with respect to the “cut distribution” in such models are unbiased. This can be useful more generally, when one wants to use different Monte Carlo techniques preserving unbiasedness, e.g. (standard) importance sampling at subsequent stages of inference. This remarkable property of the unbiased MCMC (u-MCMC) technique could also be applied beyond cut models as we now discuss.

We can extend u-MCMC to the area of Bayesian nonparametrics and the Posterior Bootstrap proposed by [Lyddon et al., 2018] to address issues arising in model misspecification. Following the notation of Algorithm 1 of [Lyddon et al., 2018], suppose one wants to obtain an unbiased estimator of $g(\tilde{\alpha})$. In the current version of Algorithm 1 this can be obtained via averaging $g(\tilde{\alpha}^{(i)})$ since it is assumed that one can draw i.i.d. samples from $\gamma^{(i)} \sim \pi(\gamma|x_{1:n})$ and $x_k^{(i)} \sim f_{\gamma(\cdot)}$, for $k = n + 1, \ldots, n + T$. In many settings, however, sampling directly from one of those distributions (or both) is impossible, e.g. when the prior for $\gamma$ is non-conjugate.

To construct an unbiased estimator of $g(\tilde{\alpha})$, we can follow the procedure described in Section 5.5 of the discussed paper for

\begin{align}
\theta_1 &:= \gamma \quad \text{and} \quad \pi_1(\gamma) \sim \pi(\gamma|x_{1:n}), \\
\theta_2 &:= (x_{(n+1):(n+T)}, w)
\end{align}

and

\[ \pi_2 ((x_{(n+1):(n+T)}, w) | \gamma) \sim f_{\gamma}(x_{n+1}) \cdot \cdots \cdot f_{\gamma}(x_{n+T}) \text{Dir} \left( \frac{w}{n}, \frac{1}{c/T}, \ldots, \frac{1}{c/T} \right), \]

\[ h(\theta_1, \theta_2) = h(\gamma, x_{(n+1):(n+T)}, w) := g \left( \arg \max_{\alpha} \sum_{j=1}^{n} w_j u (x_j, \alpha) + \sum_{j=1}^{T} w_{n+j} u (x_{n+j}, \alpha) \right). \]

(1)
Figure 1 illustrates this on a toy example from [Lyddon et al., 2018].

An interesting line of research would be developing similar methodology for adaptive MCMC where the proposal or target distributions are updated as the algorithm runs [Roberts and Rosenthal, 2009, Pompe et al., 2018]. This is particularly important in the case of two-stage estimation procedures, where after obtaining $\theta_1^{(1)}, \ldots, \theta_1^{(N_1)}$ in the second stage we run $N_1$ chains, possibly in parallel. Since each chain targets a different distribution $\pi_2(\theta_2|\theta_1^{(i)})$, there are $N_1$ transition kernels to be tuned – their mixing and consequently the efficiency of the whole algorithm could potentially be significantly improved if we could update the proposal distributions on the fly.

Figure 1: We reproduce the toy example of [Lyddon et al., 2018] (see Section 3.1 and Figure 1 of [Lyddon et al., 2018] for details) and compare these results to a sampling scheme where instead of drawing samples $\gamma_i$ from the Variational Bayes (VB) approximation, we follow the idea of u-MCMC and construct a pair of coupled Markov chains targeting this approximation. Top left panel: 95% probability contour of the true posterior and different approximations: VB and VB corrected through Posterior Bootstrap. We set the parameters of Posterior Bootstrap (PB) as $T = 100$ and $c = 1$. Top right panel: density of the coupling times of the Markov chains. Bottom panel: violin plots obtained for a function $g$ (see (1)) defined as $\tilde{\alpha}_1 \tilde{\alpha}_2$ (a product of the two coordinates) where

$\tilde{\alpha} := \arg \max_\gamma \sum_{j=1}^n w_j \log f_\gamma(x_j) + \sum_{j=1}^T w_{n+j} \log f_\gamma(x_{n+j})$.

and $\log f_\gamma$ denotes the log likelihood function. We compare results obtained with 200 samples drawn from PB and where unbiased MCMC is used at the first sampling stage, for different values of $k$ and $m$ used for defining the unbiased estimator (see Section 2.3 of the discussed paper). The parameter $k$ was each time set to 0.1$m$. Each experiment was repeated 100 times. We can see that the variance decreases dramatically as the values of $k$ and $m$ increase. It can also be observed that using Posterior Bootstrap, or PB combined with unbiased MCMC indeed corrects the VB approximation – after applying any of these procedures the results are closer to the true value of $\tilde{\alpha}_1 \tilde{\alpha}_2$ calculated for the posterior.